(12) UK Patent Application (19) GB (11) 2 295 228 (13) A

(43) Date of A Publication 22.05.1996

(21) Application No 9422688.3

(22) Date of Filing 10.11.1994

(71) Applicant(s)
Unilever Plc

(Incorporated in the United Kingdom)

Unilever House, Blackfriars, LONDON, EC4P 4BQ, United Kingdom

(72) Inventor(s)
Andrew Peter Goldsbrough
Peter Liam Jack

(74) Agent and/or Address for Service
B C Roscoe
Unilever Plc, Patent Division, Colworth House, Sharnbrook, BEDFORD, MK44 1LQ, United Kingdom

(51) INT CL[6]
C12Q 1/68

(52) UK CL (Edition O)
G1B BAC B108 B200 B220 B221 B224
U1S S1334 S1337

(56) Documents Cited
EP 0534858 A1 **WO 93/17127 A1** **WO 90/08821 A1**

(58) Field of Search
UK CL (Edition N) **G1B BAC**
INT CL[6] **C12Q 1/68**
ONLINE DATABASES: WPI,CLAIMS,DIALOG/BIOTECH

(54) Selective amplification of DNA having a plurality of sites for a specific endonuclease

(57) A process for the selective amplification of at least one part of a starting DNA which contains a plurality of restricting sites for at least one determined specific restriction endonuclease comprising:

(a) cleaving the starting DNA with a restriction endonuclease to provide a series of restriction endonuclease fragments having a region of overhang;

(b) ligation of the restriction endonuclease fragments to an adaptor molecule having a sequence of bases corresponding to the region of overhang to form a tagged restriction endonuclease fragment;

(c) contacting the tagged restriction endonuclease fragment under hybridizing conditions with at least one oligonucleotide primer;

(d) amplifying the tagged restriction endonuclease fragments hybridised with the primer by PCR; and

(e) separation of the resultant amplification products;

characterised in that the restriction endonuclease restriction site contains a sequence of unidentified bases (degeneracy), and at least either the adaptor or the primer has a specific, single sequence, is described.

GB 2 295 228 A

PROCESS FOR GENERATING DNA MARKERS

Field of Invention

This invention relates to a process for generating DNA markers for use in a number of fields including, but not limited to, plant breeding, DNA fingerprinting and most specifically to a method for detecting DNA Markers specific to the genomes of higher plants.

Background of the Invention

EP 534 858 (Keygene NV) discloses a process for the controlled amplification of at least one part of a starting DNA containing a plurality of restriction sites for a determined specific restriction endonuclease, and of which at least part of its nucleic acid is unknown. The process comprises:

(i) digesting the starting DNA with the specific restriction endonuclease or endonucleases, to fragment it into the corresponding series of restriction fragments;

(ii) ligating the restriction fragments obtained from the starting DNA with at least one double-stranded synthetic oligonucleotide (adaptor) having one end which is compatible to be ligated to one or both of the ends of the restriction fragments to thereby produce tagged restriction fragments of the starting DNA;

(iii) contacting the tagged restriction fragments under hybridizing conditions with at least 1 oligonucleotide primer;

(iv) amplifying the tagged restriction fragments hybridized with said primers by PCR or similar techniques; and

(v) identifying or recovering amplified or elongated DNA fragments as produced in step (iv);

wherein the primer includes a constant nucleotide sequence which corresponds to the nucleotides involved in the formation of the site for the restriction endonuclease and including at least part of the nucleotides present in the ligated adaptors, and a variable nucleotide sequence, located at the 3' end, which comprises a determined number of nucleotides located immediately adjacent to the last of the nucleotides involved in the restriction site for the endonuclease. Therefore in the process disclosed in EP 534 858 the restriction endonuclease has a constant nucleotide sequence.

The selection of tagged restriction fragments is determined by the number of nucleotides residing in the variable sequence part of the primer. The selectivity of the primer increases with the number of nucleotides in the variable (selected) sequence part.

It has been reported that this technology works well with small genome sizes, however problems arise when the technique is used with larger genomes (for example wheat genome).

Selection in EP 534 858 is via the primer, as described above, however such selection may not be 100% precise and polymerisation may still occur from mis-matched sites causing a background amplification. Mis-matching becomes an important problem with large genomes.

We have therefore developed an alternative process for selective restriction fragment amplification, which process has the advantage that it enables the level of mis-matching to be reduced.

Disclosure of the Invention

Accordingly the invention discloses a process for the selective amplification of at least one part of a starting DNA which contains a plurality of restriction sites for at least one determined specific restriction endonuclease comprising;

(a) cleaving the starting DNA with a restriction endonuclease to provide a series of restriction endonuclease fragments having a region of overhang;

(b) ligation of the restriction endonuclease fragments to an adaptor molecule having a sequence of bases corresponding to the region of overhang to form a tagged restriction endonuclease fragment;

(c) contacting the tagged restriction endonuclease fragment under hybridizing conditions with at least one oligonucleotide primer;

(d) amplifying the tagged restriction endonuclease fragments hybridised with the primer by PCR (polymerase chain reaction); and

(e) separation of the resultant amplification products

characterised in that the restriction endonuclease restriction site contains a sequence of unidentified bases (degeneracy), and at least either the adaptor or the primer has a specific, single sequence.

In this specification the following terms are used;

Restriction site - The nucleotide sequence recognised by the restriction endonuclease including the cleavage site. The cleavage site may be within the recognition site or

remote from it.

Recognition site - The nucleotide sequence recognised by the restriction endonuclease.

Degeneracy - The presence of a variable nucleotide sequence located within the restriction site, adaptor, or primer.

Adaptor - short double-stranded DNA molecule with a limited number of base pairs (eg. 10-30 base pairs long) which are designed in such a way that they can be ligated to the sticky end (or overhang region) in the restriction endonuclease fragment.

Rare-cutter - A restriction endonuclease whose specificity is determined by a sequence of >6 bases.

Restriction endonucleases which contain a sequence of unidentified bases within their restriction site (hereinafter described as degenerate site restriction enzymes) include at least two types of restriction endonuclease enzymes.

1. Those which have degeneracy within the recognition site. Cleavage may occur within the region of degeneracy, examples of such restriction enzymes include BstXI, BglI, SfiI, Bsa BI, BsiYI, Cac8I, DrdI EamllD5I, EcoNI, MwoI, PflMI, PshAI, XcmI XmnI and Tth111I; or cleavage may occur outside the region of degeneracy, examples of such restriction enzymes include BstEII, DdeI AlwNI, Apa BI, Dra III, and HinfI.

2. Those which cleave DNA at a specific distance (in nucleotides) from a specific recognition site, irrespective of the composition of nucleotides between the recognition site and the cleavage site. Examples of such restriction enzymes include Hga I and BcgI.

The restriction enzymes chosen must provide staggered ends, in which one of the two strands extends beyond the other (commonly known as an overhang or sticky end). Adaptors are used which have single strand extensions which are capable of annealing and ligating to the single strand extensions of the restriction fragments.

The general method of the invention involves the use of restriction endonuclease enzymes to restrict genomic DNA, ligation of synthetic oligonucleotide adaptors to genomic restriction fragments, and PCR amplification of specific subsets of the genomic restriction fragments. Selective PCR amplification of the restriction fragments can be achieved without complete prior knowledge of DNA sequences within the restriction site. This is achieved by either ligating a specific adaptor to the restriction fragments and using this and possibly further sequences within the restriction site from which to prime PCR amplification using primers which are preferably homologous to these sequences; or by ligating non-selective degenerate adaptors and using selective primers which are homologous to specific subsets of the adapted sequences; or combinations of the above two procedures.

If the adaptor incorporates a degree of degeneracy in the bases to be ligated to the digested genomic DNA, then it is essential that the PCR amplification is carried out using specific primers homologous to specific subsets of the adaptor molecule and/or to the restriction site.

Alternatively specific adaptor molecules which are homologous to subsets of the digested DNA may be ligated to the restriction endonuclease fragments. PCR amplification of these subsets is then carried out either using a degenerate (non-specific) primer or preferably using primers homologous to the adaptor and possibly extending further into a region of degeneracy within the restriction

site.

It is particularly advantageous to use both a specific adaptor molecule and a specific primer.

In a preferred process of the invention the adaptor is affinity labelled prior to it being ligated to the restriction endonuclease fragment. This allows separation of the tagged restriction endonuclease fragments from the non-tagged restriction endonuclease fragments prior to contacting with the primer. This will reduce the level of mis-matching. Suitable affinity systems include biotinylation.

Typically, the adaptors used are composed of two synthetic oligonucleotides which are in part complementary to each other, and which are usually approximately 10 to 30 nucleotides long, preferably 12 to 22 nucleotides long and which form double stranded structures when mixed together in solution. Using the enzyme ligase, the adaptors are ligated to the mixture of restriction fragments. When using a large molar excess of adaptors over restriction fragments one ensures that all restriction fragments will end up carrying adaptors at both ends.

In one embodiment of the invention, the restriction fragments carry the same adaptor at both ends and thus a single PCR primer can be used to amplify restriction fragments. In another embodiment using two (or more) different restriction endonucleases to cleave the genomic DNA, two (or more) different adaptors are ligated to the ends of the restriction fragments.

A further embodiment of the invention involves a second round of restriction endonuclease digestion (following the PCR amplification step), followed by further ligation to a second set of adaptors, and a further round of

amplification using a second set of primers. The second round of digestion would normally involve a restriction endonuclease which cuts frequently within the amplified DNA fragments.

Optionally the starting DNA may first be cleaved with a rare-cutter restriction endonuclease prior to cleaving with the degenerate site restriction enzyme. An affinity labelled adaptor specific to the rare-cutter restriction endonuclease site can then be used to further select the restriction endonuclease fragments. The rare-cutter enzyme may itself contain degeneracy.

The process of the invention provides a means to permit polymerase chain reaction amplification from selected subsets of genomic restriction fragments, allowing the detection between source samples of polymorphisms caused both by length differences (resulting from deletions, additions, inversions) which have resulted in either loss or gain of a restriction site, or changes of the nucleotide sequence in either the recognition or cleavage site of an enzyme which recognises and cleaves such sites. The invention includes methods for detecting these polymorphisms, synthetic oligonucleotides for use in the methods of the invention, applications of the methods and procedures of the invention in a number of fields including plant breeding, and DNA fingerprinting. Specifically, the methods described here provide an alternative means of identifying genomic restriction fragments which are either genetically linked to one or more particular traits, or which can provide a fingerprint of the genome under examination.

The present invention is based on the definition of novel specific methods to achieve selectivity of PCR amplification in such a way that only a small number of restriction fragments are amplified.

In general, restriction endonuclease digests of genomic DNA, and in particular of plant genomic DNA, yields very large numbers of restriction fragments, the exact number depending upon the size of the genome and the frequency of occurrence of the recognition site of the restriction endonuclease in the genome, which in turn is primarily determined by the number of nucleotides in the recognition sequence - a number typically ranging between 4 and 8. Generally the number of restriction fragments produced is too large to enable identification of individual fragments fractionated by gel electrophoresis or other fractionation methods.

We have used a number of different methods to limit the number of restriction fragments which are to be amplified. This is achieved by for example pre-selecting a subset of restriction fragments using selective adaptors or selectively amplifying using selective primers on either non pre-selected or partially pre-selected adapted genomic restriction fragments.

The basis for selection resides principally in the choice of restriction endonuclease used to digest the genomic DNA, and in the design of adaptor and primer oligonucleotides. The selective principle resides in the use of restriction endonucleases which have within their cleavage and recognition sites a number of completely degenerate nucleotides.

Selection is determined by the number of specific nucleotides in the terminal extensions of the adaptor, and/or by the number of specific nucleotides in the primer which can anneal to the non or partially selected fragments. These in turn are determined by the number of degenerate nucleotides present in the cleavage site and recognition sites of the restriction endonuclease enzyme used. The term selective bases will be used to denote both

the degenerate positions of the cleavage and recognition sequences and also the specific nucleotide sequences of the primers and adaptors to which these anneal and/or ligate.

It must be realised that amplification will only occur when the selective bases of the adaptor and/or primer are present within the cleavage and recognition sequences at both ends of the restriction fragment. When the adaptor or primer matches only one end, the amplification will be linear rather than exponential and the product will remain undetected.

It is possible to estimate the degree of selectivity obtainable by the selective adaptors or primers using the general formula 4^{2n}, where n equals the number selective bases present in the adaptor and/or primer (assuming all nucleotides are represented equally within the degenerate region). Thus, considering selective adaptors, when 1 selective base is used (the terminal base of the extension), 1 out of every 16 fragments will be capable of annealing to and ligating with the (partially) specific adaptor. Using 2 selective bases (the terminal 2 bases of the adaptor extension) 1 out of every 256 fragments will be capable of annealing to and ligating with the adaptor; using 3 selective bases (the terminal 3 bases of the adaptor extension) 1 out of every 4096 fragments will be capable of annealing to and ligating with the adaptor; using 4 selective bases (the terminal 4 bases of the adaptor extension) 1 out of every 65,536 fragments will be capable of annealing to and ligating with the adaptor; and so on. Other combinations with some completely selective and some partially selective (i.e. partially degenerate) bases give rise to different numbers of compatible ligation reactions. In a preferred embodiment the number of selective bases specifying the adaptor molecules is chosen so that the number of restriction fragments which will be amplified is limited to between 5 and 200. Although this

number can be calculated by dividing the number of restriction fragments by 4^{2n} [the number of restriction fragments can be estimated by dividing the total number of nucleotides in the genome by 4^x where x is the number of bases in the recognition sequence for the restriction enzyme used], a precise prediction is not possible because not all restriction fragments can be amplified with equal efficiency. Similar calculations can be made when the selection is based on the primer sequence for cases when both non-selective adaptors or partially selective adaptors are used. It should be pointed out that mixtures of two (or more) selective adaptors or primers can be used. This will allow selection and amplification of fragments recognised by each adaptor (or primer) and in addition those recognised by both (or all) adaptors (or primers).

The PCR amplification products obtained in accordance with the invention can be identified using standard fractionation techniques known to those skilled in the art using, for example but not limited to, agarose or acrylamide gel electrophoresis. The invention permits the number of amplified products obtained to be tuned in accordance with the resolution of the fractionation system being employed. Amplification products may be visualised directly following staining of the molecules with appropriate agents. Alternatively, the primers or nucleotides used for the PCR amplification may be labelled with radioactivity or a fluorescent chromophore, thus allowing identification of reaction products after size fractionation.

In accordance with the invention, different sets of amplified products are obtained with the different sets of selective adaptors or primers. The banding patterns identified following fractionation constitute unique and reproducible fingerprints of the genomic DNA. Such fingerprints can have several uses such as, but not

limited to, forensic typing, diagnostic identification of organisms, and the identification of species, varieties or individuals. The level of identification will be determined by the degrees of similarity and differences between the members of the group being studied. The underlying principle of the invention is that in each amplified product two nucleotide sequences are detected (the target restriction site) which are separated from each other by a given distance. In related organisms, species, varieties, races or individuals these two sequences and the relative distances separating them will be conserved to a greater or lesser degree. Hence the fingerprints obtained constitute a basis for determining the degree of sequence relationships between genomes. The fingerprints can also be used to distinguish genomes from each other.

Another particular application of the invention involves the screening and identification of restriction fragment length polymorphisms (RFLPs). Changes in the nucleotide composition of genomic DNA can often result in polymorphisms of restriction fragments: insertions or deletions affect size of the fragments containing them; nucleotide changes can result in the elimination or creation of new endonuclease target recognition sites. Restriction fragment polymorphisms of this nature can be identified directly by comparing amplified products from different genomes.

RFLPs are particularly useful for monitoring the inheritance of agronomic traits in plant breeding, in that certain DNA polymorphisms which are closely linked with specific genetic traits can be used to monitor for the presence or absence of the said trait.

A particular application of the present invention involves the detection of polymorphic amplified restriction fragments. The application involves the analysis of

amplified restriction fragment patterns obtained with different selective adaptors or primers in restriction digests of genomic DNA of closely related individuals exhibiting differences in the specific genetic trait, and the use of analysis techniques which allow correlations between the inheritance of one or more amplified restriction fragment polymorphism and the phenotype exhibited by the specific genetic traits. Once polymorphisms have been identified the polymorphic band can be purified, re-amplified, cloned, sequenced etc. and used as a marker using methods known to those skilled in the art.

The present invention provides a general method for isolating DNA markers from any genome and for using such DNA markers in all possible applications of DNA fingerprinting.

Claims

1. A process for the selective amplification of at least one part of a starting DNA which contains a plurality of restriction sites for at least one determined specific restriction endonuclease comprising:

(a) cleaving the starting DNA with a restriction endonuclease to provide a series of restriction endonuclease fragments having a region of overhang;

(b) ligation of the restriction endonuclease fragments to an adaptor molecule having a sequence of bases corresponding to the region of overhang to form a tagged restriction endonuclease fragment;

(c) contacting the tagged restriction endonuclease fragment under hybridizing conditions with at least one oligonucleotide primer;

(d) amplifying the tagged restriction endonuclease fragments hybridised with the primer by PCR; and

(e) separation of the resultant amplification products

characterised in that the restriction endonuclease restriction site contains a sequence of unidentified bases (degeneracy), and at least either the adaptor or the primer has a specific, single sequence.

2. A process according to claim 1 wherein the restriction endonuclease has degeneracy within the recognition site.

3. A process according to claim 1 wherein the restriction endonuclease cleaves DNA a fixed number of arbitrary sequence bases away from the recognition site.

4. A process according to any preceding claim wherein the adaptor molecule is degenerate in the sequence of bases corresponding to the region of overhang on the genomic fragment and the primer is specific to one of the population of adaptor molecules.

5. A process according to any one of claims 1 to 3 wherein the adaptor molecule is specific and the primer is degenerate.

6. A process according to any one of claims 1 to 3 wherein both the adaptor molecule and the primer are specific.

7. A process according to any preceding claim wherein the adaptor is affinity labelled and the sub-set of restriction endonuclease fragments which have the adaptor ligated to them are separated prior to contacting with the primer.

8. A process according to any preceding claim wherein the starting DNA is initially cleaved with a rare-cutter restriction endonuclease, which itself may contain degeneracy.

15

Patents Act 1977
Examiner's report to the Comptroller under Section 17
(The Search report)

Application number
GB 9422688.3

Relevant Technical Fields

(i) UK Cl (Ed.N) G1B (BAC)

(ii) Int Cl (Ed.6) C12Q 1/68

Search Examiner
MR C SHERRINGTON

Date of completion of Search
6 FEBRUARY 1995

Databases (see below)
(i) UK Patent Office collections of GB, EP, WO and US patent specifications.

(ii) ONLINE DATABASES: WPI, CLAIMS, DIALOG/BIOTECH

Documents considered relevant following a search in respect of Claims :-
1 to 8

Categories of documents

X: Document indicating lack of novelty or of inventive step.

Y: Document indicating lack of inventive step if combined with one or more other documents of the same category.

A: Document indicating technological background and/or state of the art.

P: Document published on or after the declared priority date but before the filing date of the present application.

E: Patent document published on or after, but with priority date earlier than, the filing date of the present application.

&: Member of the same patent family; corresponding document.

Category	Identity of document and relevant passages		Relevant to claim(s)
A	EP 0534858 A1	(KEYGENE N.V) whole document	1
A	WO 90/08821 A1	(UNIVERSITY OF MIAMI) whole document	1
A	WO 93/17127 A1	(THE STATE OF OREGON) whole document	1

Databases: The UK Patent Office database comprises classified collections of GB, EP, WO and US patent specifications as outlined periodically in the Official Journal (Patents). The on-line databases considered for search are also listed periodically in the Official Journal (Patents).